

## SPASE: The Connection Among Solar and Space Physics Data Centers

J. R. Thieman<sup>1</sup>, T. A. King<sup>2</sup>. And D. A. Roberts<sup>3</sup>

1. Code 690.1, NASA/GSFC, Greenbelt, MD, United States.
2. IGPP, 5881 Slichter Hall, UCLA, Los Angeles, CA, United States.
3. Code 672, NASA/GSFC, Greenbelt, MD, United States.

The Space Physics Archive Search and Extract (SPASE) project is an international collaboration among Heliophysics (solar and space physics) groups concerned with data acquisition and archiving. Within this community there are a variety of old and new data centers, resident archives, “virtual observatories”, etc. acquiring, holding, and distributing data. A researcher interested in finding data of value for his or her study faces a complex data environment. The SPASE group has simplified the search for data through the development of the SPASE Data Model as a common method to describe data sets in the various archives. The data model is an XML-based schema and is now in operational use. There are both positives and negatives to this approach. The advantage is the common metadata language enabling wide-ranging searches across the archives, but it is difficult to inspire the data holders to spend the time necessary to describe their data using the Model. Software tools have helped, but the main motivational factor is wide-ranging use of the standard by the community. The use is expanding, but there are still other groups who could benefit from adopting SPASE.

The SPASE Data Model is also being expanded in the sense of providing the means for more detailed description of data sets with the aim of enabling more automated ingestion and use of the data through detailed format descriptions. We will discuss the present state of SPASE usage and how we foresee development in the future. The evolution is based on a number of lessons learned – some unique to Heliophysics, but many common to the various data disciplines.

Abstract # 0082

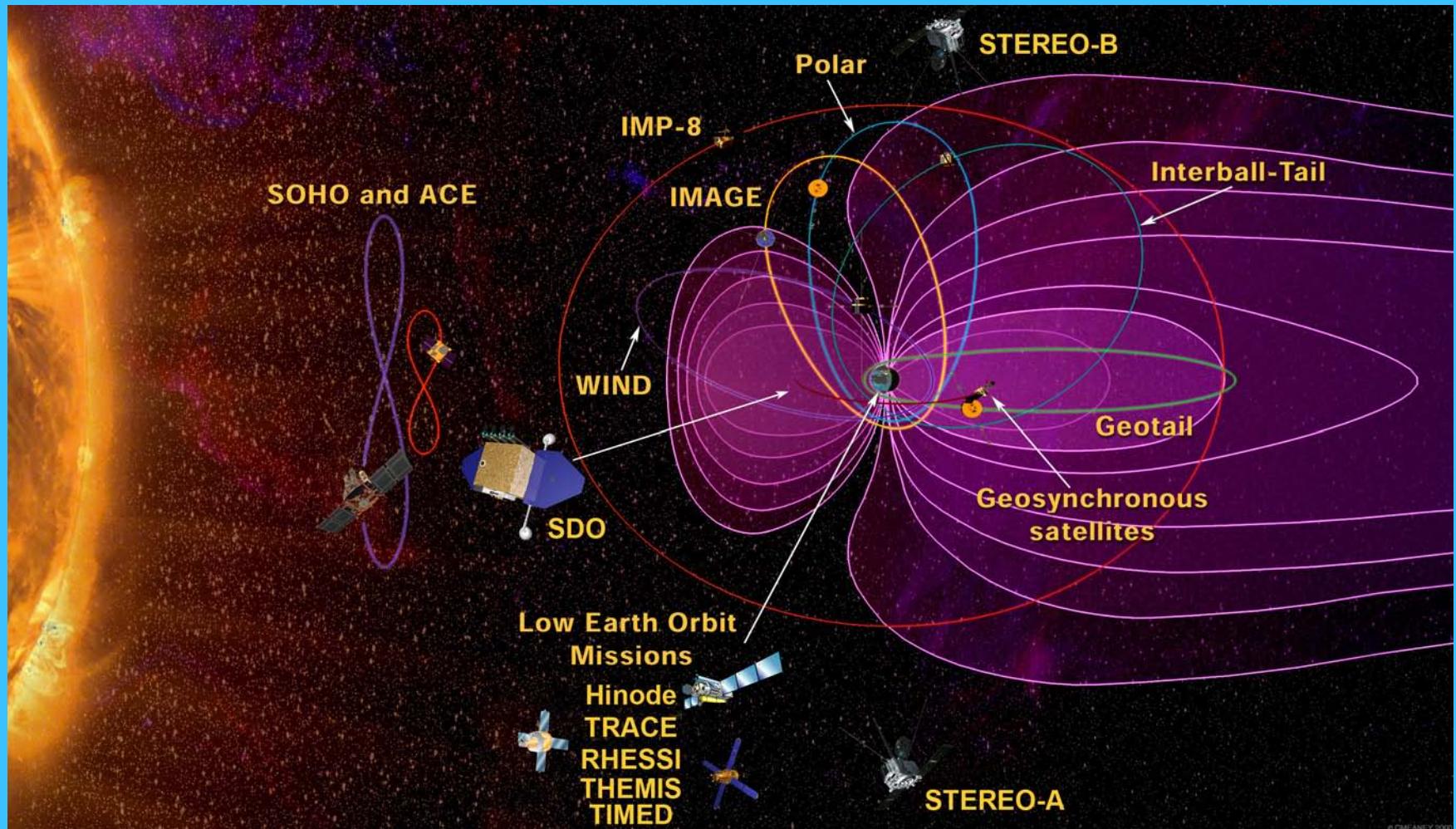
# **SPASE: The Connection Among Solar and Space Physics Data Centers**

James R. Thieman<sup>1</sup>, D. Aaron Roberts<sup>2</sup>, Todd A. King<sup>3</sup>

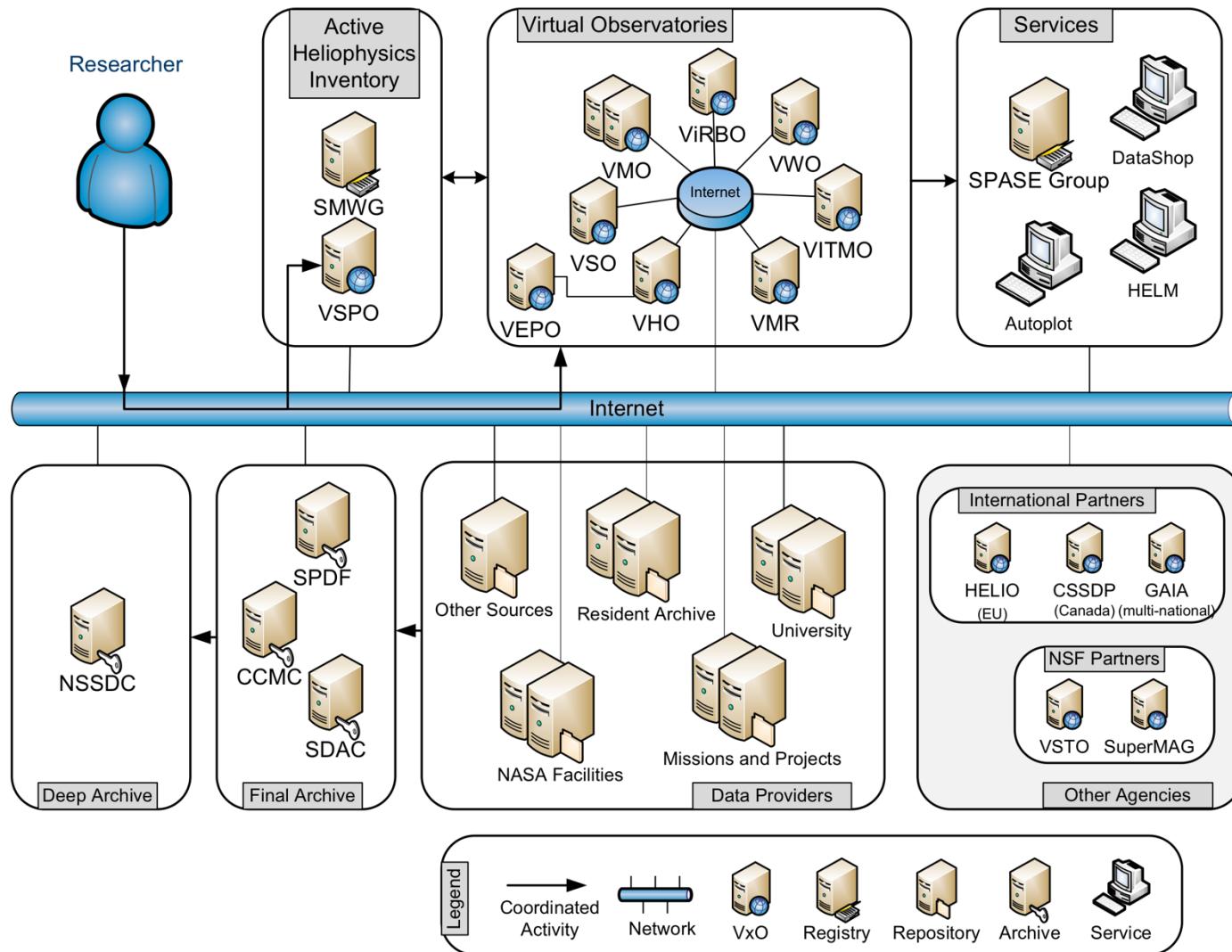
1. Code 690.1, NASA/GSFC, Greenbelt, MD, United States.
2. Code 672, NASA/GSFC, Greenbelt, MD, United States.
3. IGPP, 5881 Slichter Hall, UCLA, Los Angeles, CA, United States.

Presentation at the WDS Meeting  
Sept. 5, 2011

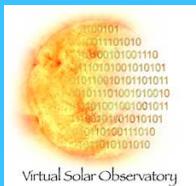
# Heliophysics (Solar and Space Physics) Great Observatory



# Heliophysics Data Environment

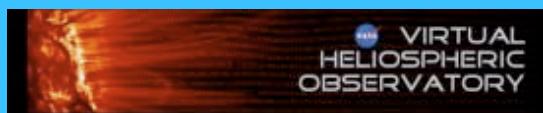


# Heliophysics Virtual Observatories (VOs)



## NASA-Funded

- VSO - Virtual Solar Observatory
- VSPO - Virtual Space Physics Observatory
- VMO - Virtual Magnetospheric Observatory
- VITMO - Virtual Ionosphere, Thermosphere, Mesosphere Observatory
- VHO - Virtual Heliophysics Observatory
- ViRBO - Virtual Radiation Belt Observatory
- VEPO - Virtual Energetic Particle Observatory
- VWO - Virtual Wave Observatory
- VMR - Virtual Model Repository



## Non-NASA-Funded

- CAA - Cluster Active Archive
- CDPP - Centre de Données de la Physique des Plasmas
- CSSDP - Canadian Space Science Data Portal
- EGSO - European Grid of Solar Observations
- GAIA - Global Auroral Imaging Access
- VSTO - Virtual Solar Terrestrial Observatory
- ??
- ??



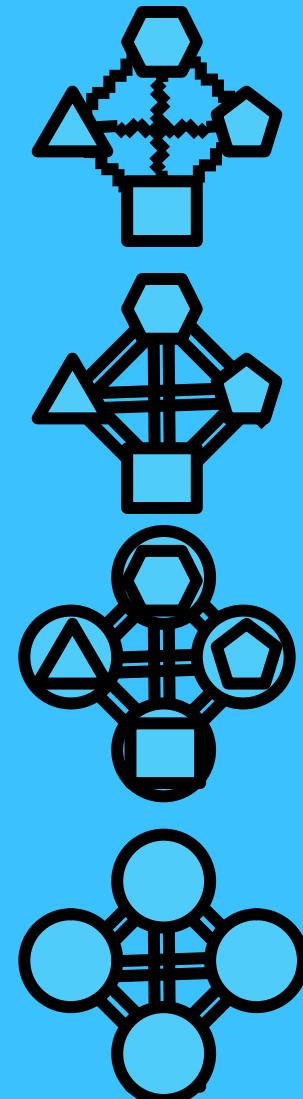
Centre de Données de la Physique des Plasmas  
Plasma Physics Data Centre

# Interoperability

**Interoperability** (among data centers, information servers, etc.) can be defined as the ability of systems in a heterogeneous network to interconnect, exchange, and use their information content in an efficient and relatively seamless way from a user point of view.

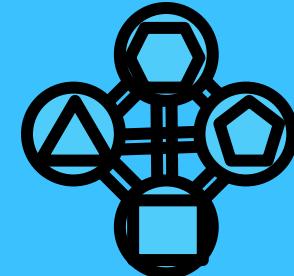
# Interoperability Levels

- 1. Basic Interconnection**
  
- 2. Information Transfer**
  
- 3. Limited Standards**
  
- 4. Uniform System**



# Level 3 Interoperability – Are We There Yet?

In Level 3 we put a common looking “shell” around the differences in the systems.



**SPASE (Space Physics Archive Search and Extract)** is an element of the “shell”.

SPASE is a **common metadata language** facilitating data search and retrieval across the Space and Solar Physics data environment.

The SPASE Group defines and maintains a standard **Data Model** for Space and Solar Physics interoperability, especially among **Heliophysics Virtual Observatories**

# ACCESS TO THE SPASE DATA MODEL



The Space Physics Archive Search and Extract (SPASE) effort is a Heliophysics community-based project with the goals of:

- Facilitating data search and retrieval across the Space and Solar Physics data environment with a common metadata language
- Defining and maintaining a standard Data Model for Space and Solar Physics interoperability, especially within the Heliophysics Data Environment
- Using the Data Model to create data set descriptions for all important Heliophysics data sets.
- Providing tools and services to assist SPASE data set description creators as well as the researchers/users
- Working with other groups for other Heliophysics data management and services coordination as needed

The Space Physics Archive Search and Extract (SPASE) effort is implemented by the SPASE Consortium which is composed of representatives of the international Heliophysics data community. The SPASE Working Group is currently the only international group supporting global data management for Solar and Space Physics.

[Learn more](#) about the SPASE group.

## Products

The SPASE Group generates three "products". First is the SPASE [Metadata Model](#) which is an information model for describing the elements of the heliophysics data environment. Second is a set of [services](#) and protocols to enable the exchange of information. Third are [tools](#) for developing and validating resource descriptions.

A complete set of [documentation](#) is also available.  
[Tutorials](#) and instructions are available at the [SPASE School](#).

**Data Model Document**

**Current Version (2.2.1)**  
Released: 2011-08-26  
[All documents](#)

[History of changes](#)

**Schema**

**Current Release (2.2.1)**  
updated: 2011-08-26  
[View all versions](#)

**Tools**

[Data Dictionary Search](#)  
[Data Model Tree](#)  
[Data Model Explorer](#)

[Registry Server](#)  
[Resource ID maker](#)  
[Editor \(web\)](#)  
[XML Validator](#)

**Services**

[SMWG Registry](#)  
[Naming Authority List](#)

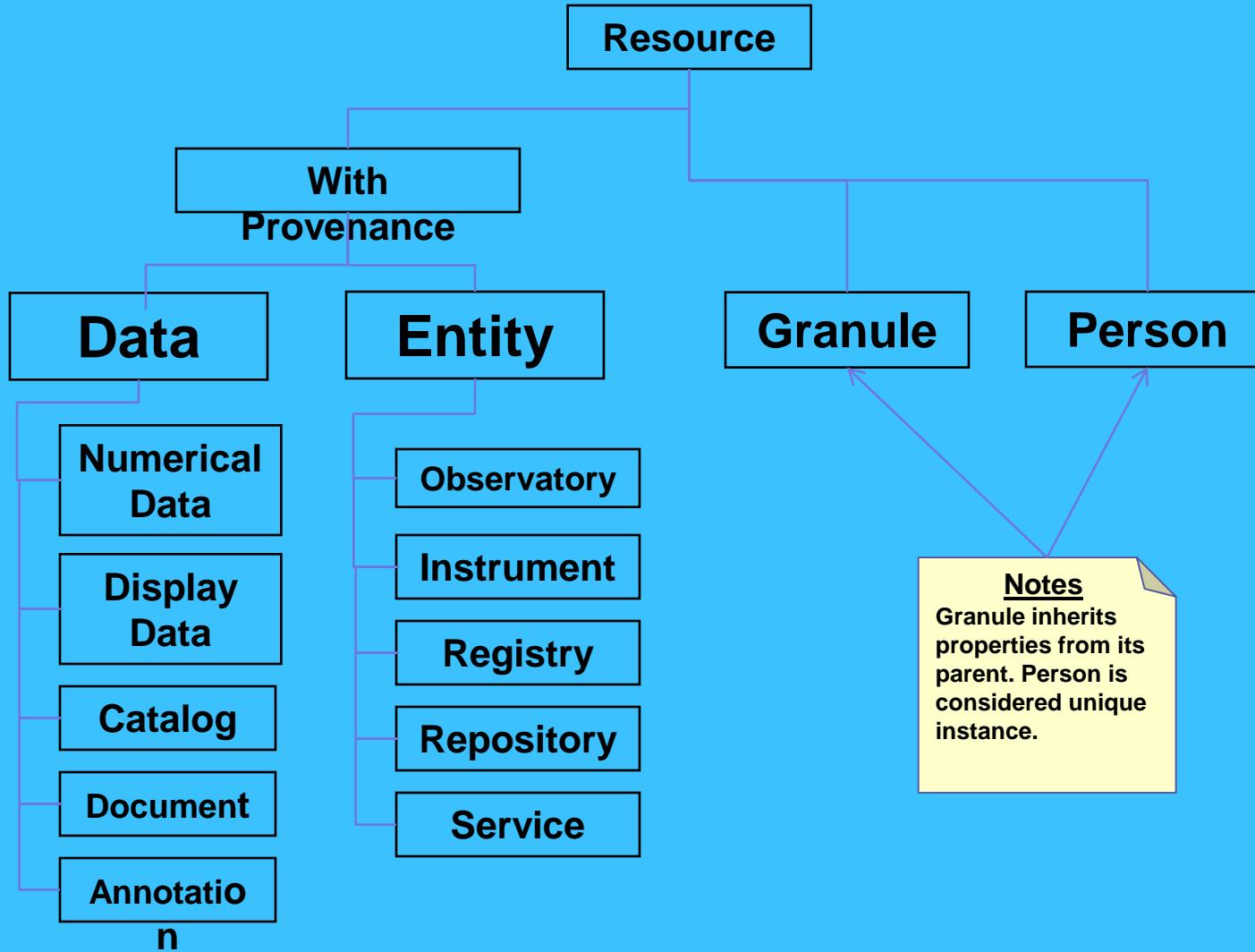
**Looking for:**

[Vintage web site](#)

**Version 2.2.1 of the Data Model is the current version and can be downloaded from here.**

<http://spase-group.org>

# SPASE Information Model



# Virtual Space Physics Observatory

The screenshot shows the homepage of the Virtual Space Physics Observatory. At the top, there's a navigation bar with links for Help, Geo Orbits, Helio Orbits, Registry, Abstracts, Weather, and Feedback. Below the navigation is a search bar with the placeholder "SEARCH NASA". To the right of the search bar is a yellow circle containing the text "SPASE inside". A yellow arrow points from this circle towards the word "FEATURES" on the right side of the slide. The main content area displays a table of search results. The columns are "# Products (& SPASE descriptions)" and "Access Links". The results are numbered 1 through 8, showing various ACE datasets (Survey Plots, CRIS L2 flux data, EPAM 1-hour and 5-minute key parameter data, and EPAM L2 and 5-min particle flux data) with their corresponding access links.

# Products (& SPASE descriptions)	Access Links
1 ACE 27-day Survey Plots	- Polar-Wind-Geotail 'gif-walk' site get data
2 ACE CRIS L2 1-day Z=3-28 flux data	- ACE Science Center - ACE/CRIS L2 data in HDF via ftp - CDAWeb - in CDF via ftp from CDAWeb get data
3 ACE CRIS L2 1-hr Z=3-28 flux data	- ACE Science Center - ACE/CRIS L2 data in HDF via ftp - CDAWeb - in CDF via ftp from CDAWeb get data
4 ACE Daily Survey Plots	- Polar-Wind-Geotail 'gif-walk' site get data
5 ACE EPAM 1-hour Key Parameter data	- in CDF via ftp from CDAWeb - CDAWeb get data
6 ACE EPAM 5-minute Key Parameter data	- in CDF via ftp from CDAWeb - CDAWeb get data
7 ACE EPAM L2 1-hour particle flux data	- ACE Science Center (ASC) - in HDF via ftp from ASC - CDAWeb - in CDF via ftp from CDAWeb get data
8 ACE EPAM L2 5-min particle flux data	- ACE Science Center (ASC) - in HDF via ftp from ASC - CDAWeb - in CDF via ftp from CDAWeb get data

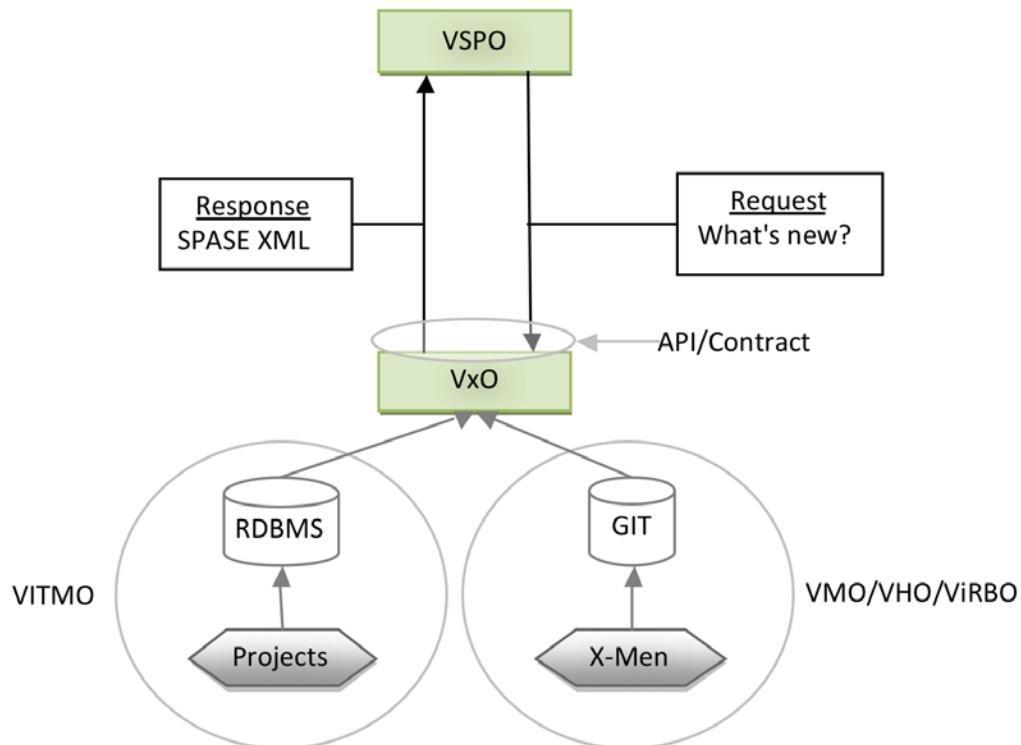
## FEATURES

- Contains all SPASE data set descriptions
- Uses SPASE terminology for search keywords
- Provides data access through “get data” links
- Queries easily modified
- Can ingest XML-based SPASE descriptions

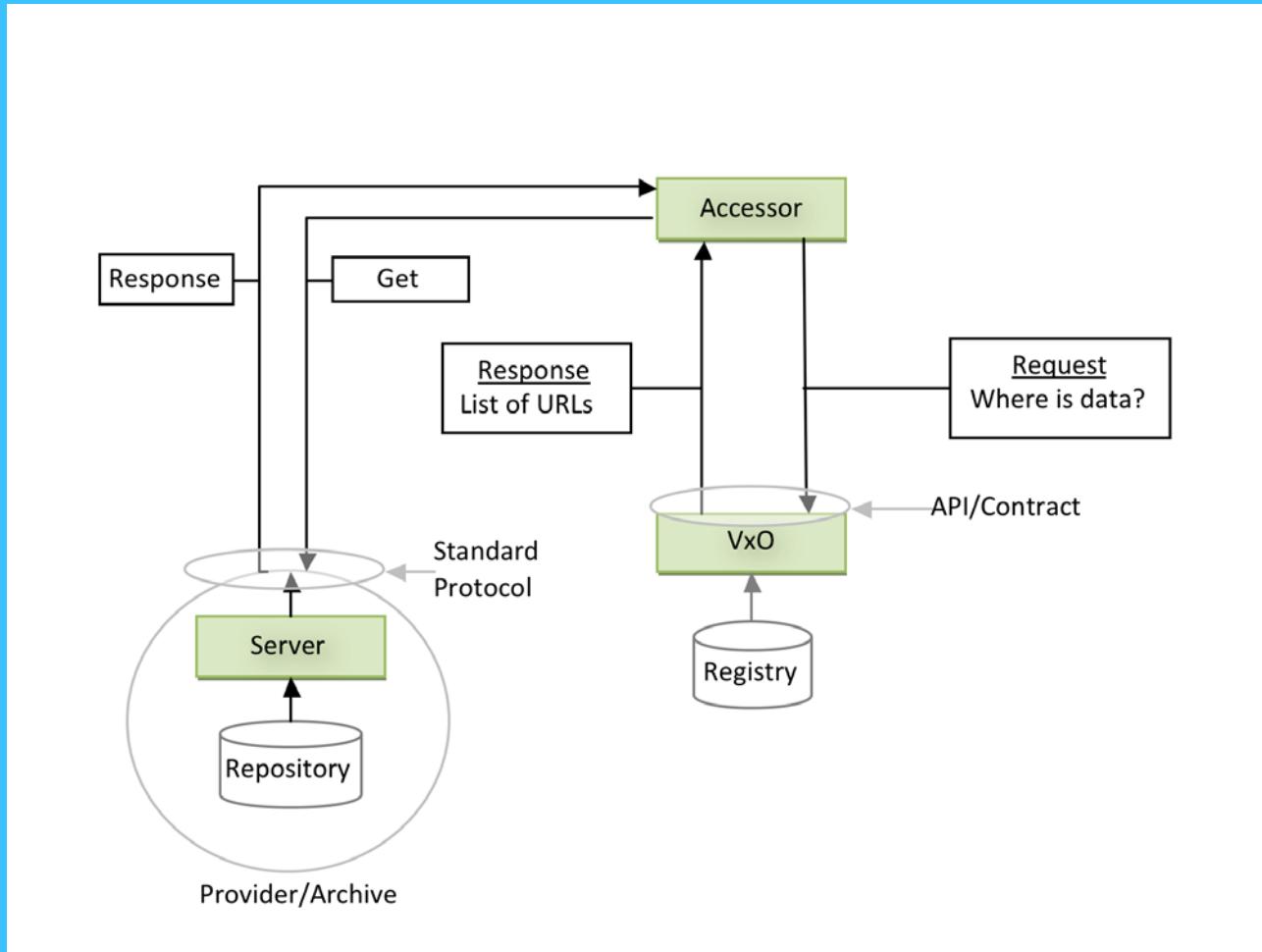
<http://vspo.gsfc.nasa.gov/websearch/dispatcher>

or google “VSPO”

# Harvesting



# Data Extract



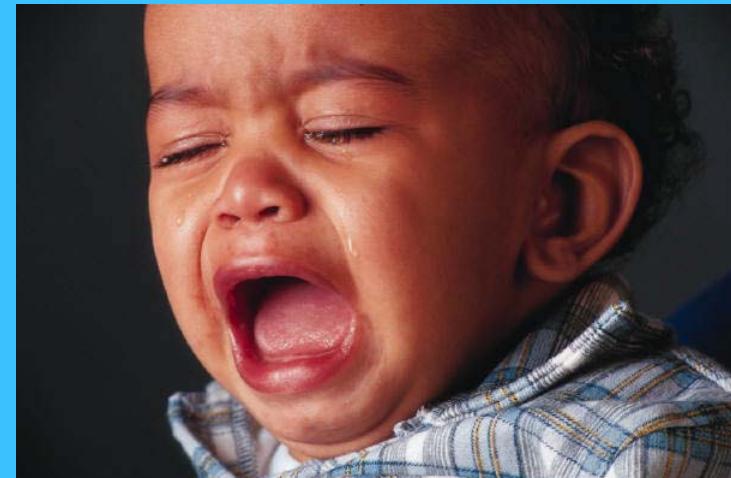
# The Human Factor

**The use of SPASE depends, of course, on the willingness of the data holders to describe their data via SPASE.**

**It is human nature to avoid writing data descriptions.**

**Simple, high-level descriptions are often sufficient for the purpose.**

**More detailed descriptions are better for the researcher, so many tools have been created to make data descriptions and the use of them easier to do.**



# Application Tools

*Tools for working with SPASE metadata and the SPASE framework.*

## Generator

Create SPASE descriptions using external sources of information

[Ruleset Description Generator](#)

## Editor

### Web-based Editors

[Web Editor](#)

### Standalone Editors

[SPASE Assistant](#)

### Editors with Database Storage

[Web+DB Editor](#)

## Validator

Determines compliance with a version of the SPASE data model.

[XML Validate](#)

## Parser

Convert SPASE XML to internal structures

[Parser](#)

## Harvester

Extracts information from SPASE resource descriptions (or registries)

[SPASE Registry Server](#)

[SPASE Database Query](#)

## Wrapper

Converts or embeds SPASE metadata in other descriptions or forms (i.e., OAI)

[Data Dictionary Lookup](#)

[SPASE-to-OAI mapping](#)

## Correlator

Divide an XML document into individual resource descriptions into a well organized file system

[Correlator](#)

## Refcheck

Determine the validity of all references in a resource descriptions. Checks Resource IDs and URL

[Refcheck](#)

*There are additional tools in development:*

[SPASE Query Language](#)

[Java-to-XML Binding Mechanism \(JAXB\)](#)

[SPASE Guidelines Document](#)

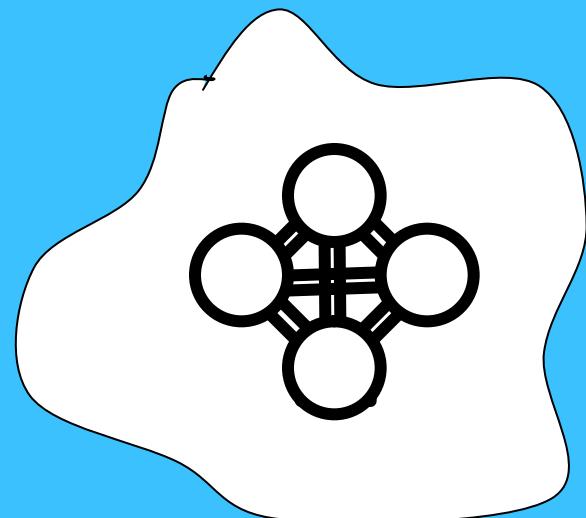
# Is Level 4 in the Clouds?

**Level 4 Interoperability –  
indistinguishable from a single  
all-encompassing system.**

**Is this achievable?**

**Is this “the cloud”?**

**Is this desirable?**



# Summary

- The Heliophysics Data Environment is historically diverse and widespread, but is being unified
- SPASE provides a standard metadata approach toward unification
- Creation of SPASE data descriptions is key to progress
- Utility of SPASE tools can greatly influence success
- Will cloud computing be a step toward further unification?